Massive MIMO

Luca Sanguinetti

Abstract The aim of this chapter is to provide an entry point to Massive MIMO, as well as an up-to-date survey of the state-of-the-art results in spectral efficiency, that will guide the evolution of Massive MIMO in the years to come. Particularly, it builds upon the book [1] that is freely accessible in PDF from the website www.massivemimobook.com.

1 Introduction

The area throughput is highly relevant in contemporary and future cellular networks. It is measured in bit/s/km^2 and determined by three key factors:

$$\text{Area throughput [bit/s/km}^2\text{]} = \frac{B\text{ [Hz]} \times \text{Spectral efficiency [bit/s/Hz/cell]}}{\text{Cell size [km}^2\text{/cell]}} \quad (1)$$

where B is the bandwidth, and the spectral efficiency (SE) is the amount of information that can be transferred per second over one Hz of bandwidth. Despite inherently dependent, the three components can be treated as independent as a first-order approximation. Consequently, the area throughput can be improved by:

1. Allocating more bandwidth;
2. Densifying the network by deploying more base stations (BSs);
3. Improving the SE per cell.

The traditional way to increase the area throughput in cellular networks is to allocate more bandwidth and to deploy more BSs (i.e., to reduce cell size). Consequently, contemporary networks are already densely deployed and the bandwidth

Prof. L. Sanguinetti, Ph.D. (IEEE Senior Member)
University of Pisa, Dipartimento di Ingegneria dell'Informazione, Italy
e-mail: luca.sanguinetti@unipi.it

left in the sub-6 GHz bands (that are attractive for wide-area coverage) is scarse [2]. In contrast, the growth in SE has been rather modest over the last years.

Nowadays, one of the most promising wireless technology to improve the SE is Massive MIMO (multiple-input multiple-output) [3, 4, 5]. This physical-layer technology was introduced by Thomas Marzetta's seminal paper from 2010 [7] and equips each BS with an array of many antennas. This allows for coherent multiuser MIMO transmission where tens of user equipments (UEs) can be served on each time-frequency resource by spatial multiplexing, in both the uplink (UL) and downlink (DL) of each cell. The multiplexing gain allows to improve the SE per cell by orders-of-magnitude [1]. In [7], Marzetta showed that a system with an extremely large number of BS antennas should operate in time-division duplex (TDD) mode and exploit channel reciprocity to acquire all the necessary channel state information (CSI) from a finite number of UL pilot signals. Massive MIMO has since then gradually changed from a controversial theoretical concept to a mainstream technology that has found its way into the 5G standard [8].

2 What is Massive MIMO?

Although Massive MIMO is widely considered as one of the key technologies to provide high SE, it lacks a concise and universal definition. The telecom industry treats it as multiuser MIMO with many BS antennas, in order to quickly reach the market. To demonstrate how such a physical-layer technology can improve the SE by orders-of-magnitude, in this chapter we will consider a stricter definition.

Definition 1 A Massive MIMO cellular network consists of:

- $L \geq 2$ cells operating according to a synchronous TDD protocol;
- BSs equipped with $M \geq 64$ antennas with fully digital transceiver chains;
- Linear combining and precoding schemes capable of spatially multiplexing $K \geq 8$ UEs per cell;
- More BS antennas than active UEs: $M/K > 1$.

This definition is in line with the canonical form of Massive MIMO for sub-6 GHz bands in [1] and includes Marzetta's setup from [7] as a special case as $M \to \infty$. It is also in line with real-time Massive MIMO testbeds [9] and field trials [10]. However, there are important research efforts that deviate from it. Finding an efficient frequency-division duplexing (FDD) protocol for Massive MIMO is highly desirable, since there are vast amounts of spectrum reserved for FDD operation. However, the estimation and feedback overhead of FDD operation in mobile scenarios is prohibitive, unless something is done to reduce it. The predominant approach is to

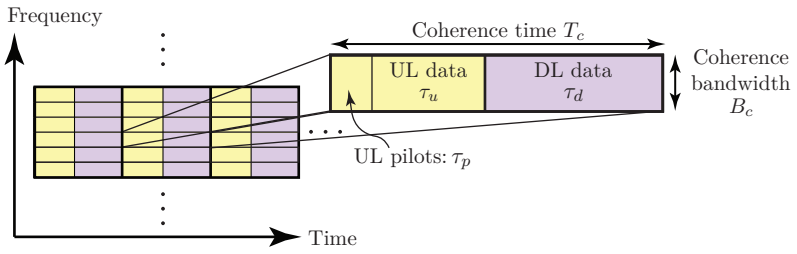

Fig. 1 The time-frequency plane is divided into coherence blocks over which the channel response is time-invariant and frequency-flat. The τ_c samples of each block are used in a TDD fashion for UL pilots, UL data, and DL data.

assume that there is some kind of channel sparsity that can be utilized to reduce the channel estimation and feedback overhead. This line of research is quite rich (e.g., [11, 12]) but the underlying sparsity hypothesis has not been proved experimentally at sub-6 Ghz frequencies [13]. Another deviation from the canonical form is the use of the mmWave band, which refers to the frequency range from 30 – 300 GHz. Although there is a vast literature that use the "Massive MIMO" term for both sub-6 GHz and mmWave applications, this can be very confusing because the multiantenna technology has rather different characteristics in these two applications. Massive MIMO at sub-6 GHz spectrum can increase the efficiency of highly loaded cells, by upgrading the technology at existing BSs. In contrast, the huge available bandwidth in mmWave band can be utilized for high-capacity services, but only over short distances due to the severe pathloss and high noise power.

2.1 The Transmission Protocol

Massive MIMO can only make efficient use of antennas if each BS knows the UE channels. For this purpose, a TDD protocol is used where the UL and DL transmissions fit into the channel *coherence block*, as illustrated in Fig. 1. This represents the time-frequency block in which the channels can be approximated as time-invariant and flat-fading. The number of samples τ_c in a coherence block is determined by the coherence time T_c and coherence bandwidth B_c of channels. Typical values for τ_c ranges from hundreds (with high-mobility and high-channel dispersion) to thousands of samples (with low-mobility and low-channel dispersion) [1, Remark 2.1]. As shown in Fig. 1, the τ_c samples are used[1] for three purposes: 1) τ_p samples for UL pilots; 2) τ_u samples for UL reception; 3) τ_d samples for DL transmission.

The UL pilots enable the BS to estimate UE channels. Since the TDD protocol is matched to the coherence block, the UL and DL channels can be considered reciprocal[2] and the BS can use channel estimates for UL reception and DL transmission.

[1] Variations of the TDD protocol above exist, for which we refer to [14, 15].

[2] The radio frequency propagation channels are reciprocal by nature, but the end-to-end channels are also affected by the transceiver hardware. We refer to [17] for reciprocity calibration algorithms.

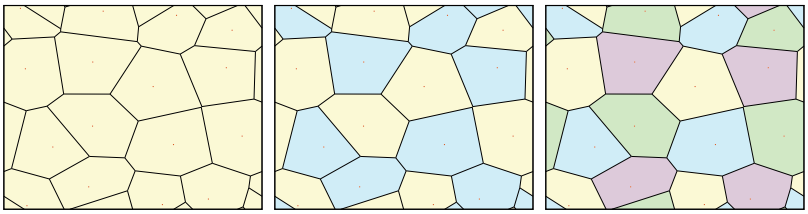

Fig. 2 Illustration of a cellular network with a pilot reuse factor of $f = 1, 2$ or 4. Each group is indicated with a distinct color and uses a disjunct set of pilots.

A common assumption is $\tau_p = Kf \leq \tau_c$, so that a pilot book $\boldsymbol{\Phi} \in \mathbb{C}^{\tau_p \times \tau_p}$ of Kf mutually orthogonal pilot sequences can be created, for example, by using the columns of a Walsh-Hadamard matrix [1]. The pilot book is divided into f groups with $K = \tau_p/f$ pilots each, where the integer f is called the *pilot reuse factor*. Each cell is associated with one of these disjunct groups, according to a predefined pilot reuse pattern, as illustrated in Fig. 2. Hence, the K UEs in a cell use mutually orthogonal pilots. The pilot associated with UE k in cell j is denoted by $\boldsymbol{\phi}_{jk} \in \mathbb{C}^{\tau_p}$ and is transmitted with power ρ_{ul}. This implies $\boldsymbol{\phi}_{jk}^{\mathrm{T}} \boldsymbol{\phi}_{jk}^{*} = \tau_p \rho_{\text{ul}}$. For simplicity, we assume that UE k in each cell uses the same pilot, and call $\mathcal{P}_j \subset \{1, \ldots, L\}$ the group of cells that utilize the same pilot set of cell j.

2.2 Spatially Correlated Channels

We denote by $\mathbf{h}_{lk}^{j} \in \mathbb{C}^{M}$ the channel between UE k in cell l and BS j. In the Massive MIMO literature, this channel is often modeled as $\mathbf{h}_{lk}^{j} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_M, \beta_{lk}^{j} \mathbf{I}_M)$ [3] where the Gaussian distribution accounts for the random small-scale fading, while β_{lk}^{j} describes the macroscopic large-scale fading. This channel model is called *uncorrelated Rayleigh fading*, and makes the analysis as well as the optimization of Massive MIMO tractable. However, it is practically questionable. Indeed, propagation channels are spatially correlated, as seen from measurement campaigns [16] and by the fact that uncorrelation only appears under extreme physical circumstances. [1, Sec. 2.2]. A tractable way to model spatially correlated channels is the *correlated Rayleigh fading* model:

$$\mathbf{h}_{lk}^{j} \sim \mathcal{N}_{\mathbb{C}} \left(\mathbf{0}_M, \mathbf{R}_{lk}^{j} \right) \tag{2}$$

where $\mathbf{R}_{lk}^{j} \in \mathbb{C}^{M \times M}$ is the spatial correlation matrix, assumed to be known at the BS. As with uncorrelated Rayleigh fading, the Gaussian distribution models the small-scale fading variations, while $\mathbf{R}_{lk}^{j}$ describes the macroscopic propagation characteristics. The normalized trace $\beta_{lk}^{j} = \frac{1}{M} \mathrm{tr}(\mathbf{R}_{lk}^{j})$ is the average channel gain from an antenna at BS j to UE k in cell l. Uncorrelated Rayleigh fading with $\mathbf{R}_{lk}^{j} = \beta_{lk}^{j} \mathbf{I}_M$ is a special case of this model, but $\mathbf{R}_{lk}^{j}$ is in general not diagonal. The eigenstructure of $\mathbf{R}_{lk}^{j}$ determines the spatial channel correlation of $\mathbf{h}_{lk}^{j}$; that is, which spatial directions are statistically more likely to contain strong signal components

than others. Strong spatial correlation is characterized by large eigenvalue variations. We refer to [1, Sec. 7.3] for a detailed description of how to generate $\mathbf{R}_{lk}^{j}$.

2.3 Channel Estimation

As said earlier, to make efficient use of its antennas, a given BS j needs to learn channels. Since the pilots are transmitted synchronously in all cells, BS j can use them to estimate the channels from its own UEs and the channels from UEs in other cells. The latter estimates can be used for inter-cell interference suppression.

Channel estimates at BS j are obtained from the received pilot signal $\mathbf{Y}_{j}^{p} \in \mathbb{C}^{M \times \tau_p}$, which is given by

$$\mathbf{Y}_{j}^{p} = \underbrace{\sum_{i=1}^{K} \mathbf{h}_{ji}^{j} \boldsymbol{\phi}_{ji}^{\mathrm{T}}}_{\text{Desired pilots}} + \underbrace{\sum_{l=1, l \neq j}^{L} \sum_{i=1}^{K} \mathbf{h}_{li}^{j} \boldsymbol{\phi}_{li}^{\mathrm{T}}}_{\text{Inter-cell pilots}} + \underbrace{\mathbf{N}_{j}^{p}}_{\text{Noise}} \tag{3}$$

where $\mathbf{N}_{j}^{p} \in \mathbb{C}^{M \times \tau_p}$ is thermal noise with i.i.d. elements distributed as $\mathcal{N}_{\mathbb{C}}(0, \sigma_{\mathrm{ul}}^{2})$. If the statistics are known, the minimum mean-squared error (MMSE) estimator of $\mathbf{h}_{li}^{j}$ can be computed as follows.

Theorem 1 *The MMSE estimate of* $\mathbf{h}_{li}^{j}$ *is*

$$\hat{\mathbf{h}}_{li}^{j} = \mathbf{R}_{li}^{j} \left(\mathbf{Q}_{li}^{j}\right)^{-1} \left(\frac{1}{\tau_p \sqrt{\rho_{\mathrm{ul}}}} \mathbf{Y}_{j}^{p} \boldsymbol{\phi}_{li}^{*} \right) \tag{4}$$

where $\mathbf{Y}_{j}^{p}$ *is given in* (3) *and*

$$\mathbf{Q}_{li}^{j} = \sum_{l' \in \mathcal{P}_l} \mathbf{R}_{l'i}^{j} + \frac{1}{\tau_p} \frac{\sigma_{\mathrm{ul}}^{2}}{\rho_{\mathrm{ul}}} \mathbf{I}_{M}. \tag{5}$$

The estimation error $\tilde{\mathbf{h}}_{li}^{j} = \mathbf{h}_{li}^{j} - \hat{\mathbf{h}}_{li}^{j}$ *is independent of* $\hat{\mathbf{h}}_{li}^{j}$ *and has correlation matrix* $\mathbf{C}_{li}^{j} = \mathbb{E}\{\tilde{\mathbf{h}}_{li}^{j} (\tilde{\mathbf{h}}_{li}^{j})^{\mathrm{H}}\} = \mathbf{R}_{li}^{j} - \boldsymbol{\Phi}_{li}^{j}$ *with* $\boldsymbol{\Phi}_{li}^{j} = \mathbf{R}_{li}^{j} (\mathbf{Q}_{li}^{j})^{-1} \mathbf{R}_{li}^{j}$.

The normalized mean-squared error (NMSE) given by

$$\frac{\mathbb{E}\{\|\mathbf{h}_{li}^{j} - \hat{\mathbf{h}}_{li}^{j}\|^2\}}{\mathbb{E}\{\|\mathbf{h}_{li}^{j}\|^2\}} = \frac{\mathrm{tr}(\mathbf{C}_{li}^{j})}{\mathrm{tr}(\mathbf{R}_{li}^{j})} = 1 - \frac{\mathrm{tr}(\mathbf{R}_{li}^{j} (\mathbf{Q}_{li}^{j})^{-1} \mathbf{R}_{li}^{j})}{\mathrm{tr}(\mathbf{R}_{li}^{j})} \tag{6}$$

shows that the interference generated by the pilot-sharing UEs in $\mathcal{P}_l$, which enters into $\mathbf{Q}_{li}^{j}$ in (5), increases it and thus reduces the channel estimation quality. This "pilot interference" is called *pilot contamination* and behaves differently from noise; it not only reduces the estimation quality, but impacts also the SE since (4) contains the channels of pilot-sharing UEs [1, Sec. 4.4.2]. If $\mathbf{R}_{li}^{j} \mathbf{R}_{l'i}^{j} = \mathbf{0}_{M}$, the NMSE in

(6) is completely unaffected by UE k in cell l [1, Sec. 3.3.2]. Therefore, in theory, pilot contamination between two UEs can be completely avoided if their correlation matrices are spatially orthogonal. While this condition is unlikely to hold in practice [1, 5], assigning pilots such that $\mathrm{tr}(\mathbf{R}_{li}^{j}\mathbf{R}_{l'i}^{j})$ is rather small is a good rule-of-thumb [18].

3 Spectral Efficiency

We analyze the achievable SE, focusing for simplicity only on the UL. The data signal from UE k in cell j is denoted by $s_{jk} \sim \mathcal{N}_{\mathbb{C}}(0, \rho_{\mathrm{ul}})$, with ρ_{ul} being the transmit power. To detect s_{jk}, BS j selects the combining vector $\mathbf{v}_{jk} \in \mathbb{C}^M$, which is multiplied with the received signal $\mathbf{y}_j \in \mathbb{C}^M$ at BS j. This yields

$$\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{y}_j = \mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{jk}^{j}s_{jk} + \underbrace{\sum_{i=1,i\neq k}^{K}\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{ji}^{j}s_{ji}}_{\text{Intra-cell interference}} + \underbrace{\sum_{l=1,l\neq j}^{L}\sum_{i=1}^{K}\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{h}_{li}^{j}s_{li}}_{\text{Inter-cell interference}} + \underbrace{\mathbf{v}_{jk}^{\mathrm{H}}\mathbf{n}_j}_{\text{Noise}} \tag{7}$$

where $\mathbf{n}_j \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_M, \sigma_{\mathrm{ul}}^2\mathbf{I}_M)$ is independent noise. An *achievable SE* is any number that is below the capacity. While the classical "Shannon formula" cannot be applied when the receiver has imperfect CSI, there exist well-established capacity lower bounds that can be used [1, 3].

Theorem 2 *If the MMSE channel estimation is used, an UL SE of UE k in cell j is*

$$\mathsf{SE}_{jk}^{\mathrm{ul}} = \frac{\tau_u}{\tau_c}\mathbb{E}\left\{\log_2\left(1+\gamma_{jk}^{\mathrm{ul}}\right)\right\} \quad \textit{[bit/s/Hz]} \tag{8}$$

with τ_u/τ_c accounting for the fraction of samples used for UL data and

$$\gamma_{jk}^{\mathrm{ul}} = \frac{|\mathbf{v}_{jk}^{\mathrm{H}}\hat{\mathbf{h}}_{jk}^{j}|^2}{\mathbf{v}_{jk}^{\mathrm{H}}\left(\sum\limits_{l=1,l\neq j}^{L}\sum\limits_{i=1}^{K}\hat{\mathbf{h}}_{li}^{j}(\hat{\mathbf{h}}_{li}^{j})^{\mathrm{H}} + \sum\limits_{i=1,i\neq k}^{K}\hat{\mathbf{h}}_{jk}^{j}(\hat{\mathbf{h}}_{jk}^{j})^{\mathrm{H}} + \mathbf{Z}_j\right)\mathbf{v}_{jk}} \tag{9}$$

where $\mathbf{Z}_j = \sum_{l=1}^{L}\sum_{i=1}^{K}\left(\mathbf{R}_{li}^{j} - \mathbf{\Phi}_{li}^{j}\right) + \frac{\sigma_{\mathrm{ul}}^2}{\rho_{\mathrm{ul}}}\mathbf{I}_M$.

The bound in (8) has been analyzed in a number of articles that consider heuristic linear detectors [3]. Two popular choices for $\mathbf{v}_{jk}$ are maximum-ratio (MR) and zero-forcing (ZF) combining:

$$\mathbf{V}_j \triangleq \left[\mathbf{v}_{j1}\ \ldots\ \mathbf{v}_{jK}\right] = \begin{cases}\mathbf{V}_j^{\mathrm{MR}} = \widehat{\mathbf{H}}_j^{j} & \text{with MR combining}\\ \mathbf{V}_j^{\mathrm{ZF}} = \widehat{\mathbf{H}}_j^{j}\left((\widehat{\mathbf{H}}_j^{j})^{\mathrm{H}}\widehat{\mathbf{H}}_j^{j}\right)^{-1} & \text{with ZF combining}\end{cases} \tag{10}$$

with $\widehat{\mathbf{H}}_j^{j} = [\widehat{\mathbf{h}}_{j1}^{j}\ \ldots\ \widehat{\mathbf{h}}_{jK}^{j}] \in \mathbb{C}^{M\times K}$ containing the estimates of intra-cell channels in cell j. In a single-cell scenario with perfect CSI, MR and ZF are asymptotically

Table 1 Network parameters for SE evaluation. The asymmetric network (with wrap-around) in Fig. 2 is considered.

Parameter	Value
Cell area	$1\,\text{km} \times 1\,\text{km}$
Number of cells and UEs per cell	$L = 16$, $K = 10$
UL noise power and UL transmit power	$\sigma^2_{\text{ul}} = -94\,\text{dBm}$, $\rho_{\text{ul}} = 20\,\text{dBm}$
Samples per coherence block	$\tau_c = 200$
Pilot reuse factor	$f = 1, 2$ or 4
Distance between UE k in cell l and BS j	d^j_{lk}
Large-scale fading coefficient for the channel between UE k in cell l and BS j	$\beta^j_{lk} = -148.1 - 37.6 \log_{10}\left(\frac{d^j_{lk}}{1\,\text{km}}\right) + F^j_{lk}\,\text{dB}$
Shadow fading between UE k in cell l and BS j	$F^j_{lk} \sim \mathcal{N}(0, 10)$

optimal at low and high SNRs [1, Sec. 4.1.1]. However, in a multicellular network with imperfect CSI and pilot contamination, they are both suboptimal. Instead of resorting to heuristics, we notice that the SINR in (8) has the form of a generalized Rayleigh quotient. Hence, the maximum is achieved by [5]:

$$\mathbf{V}_j^{\text{M-MMSE}} \triangleq \left[\mathbf{v}_{j1} \ \ldots \ \mathbf{v}_{jK}\right] = \left(\sum_{l=1}^{L} \widehat{\mathbf{H}}_l^j \big(\widehat{\mathbf{H}}_l^j\big)^{\text{H}} + \mathbf{Z}_j\right)^{-1} \widehat{\mathbf{H}}_j^j. \tag{11}$$

This optimal combining scheme also minimizes the conditional MSE $\mathbb{E}\{|s_{jk} - \mathbf{v}_{jk}^{\text{H}}\mathbf{y}_j|^2 \mid \{\widehat{\mathbf{H}}_l^j\}\}$. It was introduced in [5, 19] and called multicell MMSE (M-MMSE) combining. The "multicell" notion was used to differentiate it from the single-cell MMSE (S-MMSE) combining scheme [20], which is widely used in the Massive MIMO literature and given by:

$$\mathbf{V}_j^{\text{S-MMSE}} \triangleq \left[\mathbf{v}_{j1} \ \ldots \ \mathbf{v}_{jK}\right] = \left(\widehat{\mathbf{H}}_j^j \big(\widehat{\mathbf{H}}_j^j\big)^{\text{H}} + \overline{\mathbf{Z}}_j\right)^{-1} \widehat{\mathbf{H}}_j^j \tag{12}$$

with $\overline{\mathbf{Z}}_j = \sum_{i=1}^{K} \left(\mathbf{R}_{ji}^j - \boldsymbol{\Phi}_{ji}^j\right) + \sum_{l=1, l\neq j}^{L} \sum_{i=1}^{K} \mathbf{R}_{li}^j + \frac{\sigma^2_{\text{ul}}}{\rho_{\text{ul}}} \mathbf{I}_M$. The main difference from (11) is that only channel estimates in the own cell are computed and utilized in S-MMSE, while $\hat{\mathbf{h}}_{jli}\hat{\mathbf{h}}_{jli}^{\text{H}} - \boldsymbol{\Phi}_{jli}$ is replaced with its average (i.e., zero) for all UEs in other cells.

3.1 Performance Evaluation

To quantify the SE that can be achieved in Massive MIMO, we now consider the network setup in Fig. 2 with the parameters given in Table 1. Each BS is equipped with a uniform linear array with half-wavelength antenna spacing. Each channel consists of $S = 6$ scattering clusters, which are modeled by the Gaussian local scattering model [1, Sec. 2.6]. Hence, the (m_1, m_2)th element of $\mathbf{R}_{li}^j$ is given by

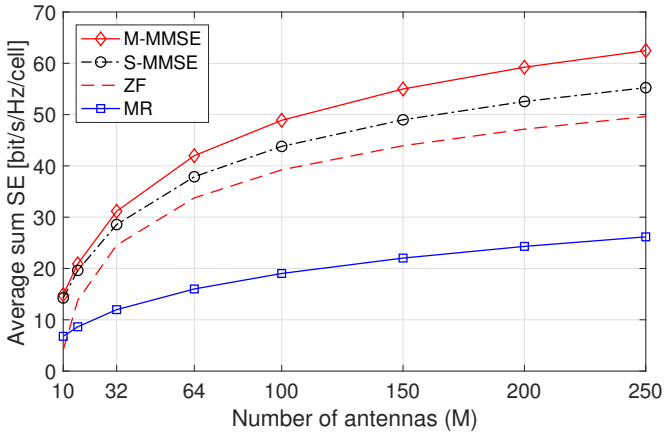

Fig. 3 Average UL sum SE as a function of M with different combining schemes with the correlated Rayleigh fading model in (13), by using the network setup in Fig. 2 with pilot reuse factor $f = 1$.

$$\left[\mathbf{R}_{li}^{j}\right]_{m_1,m_2} = \beta_{li}^{j} 10^{\frac{f_{m_1}+f_{m_2}}{10}} \frac{1}{S}\sum_{s=1}^{S} e^{\mathrm{j}\pi(m_1-m_2)\sin(\varphi_{li,s}^{j})} e^{-\frac{\sigma_{\varphi}^{2}}{2}\left(\pi(m_1-m_2)\cos(\varphi_{li,s}^{j})\right)^{2}} \quad (13)$$

where β_{li}^{j} is reported in Table 1 and $f_m \sim \mathcal{N}(0,\sigma_f^2)$ represents i.i.d. log-normal channel gain variations with $\sigma_f = 2$, which model the gain variations observed from measurements in [13]. Let φ_{li}^{j} be the geographical angle to UE i in cell l as seen from BS j. Cluster s is characterized by the randomly generated nominal angle-of-arrival $\varphi_{li,s}^{j} \sim \mathcal{U}[\varphi_{li}^{j} - 40°, \varphi_{li}^{j} + 40°]$ and the angles of the multipath components are Gaussian distributed around the nominal angle with standard deviation $\sigma_{\varphi}^{2} = 5°$.

Fig. 3 shows the average UL sum SE for the pilot reuse factor $f = 1$. M-MMSE provides the highest SE, which passes from 14.89 bit/s/Hz to 62.47 bit/s/Hz as M increases. The suboptimal schemes are quite competitive when M is small, but in the Massive MIMO regime of $M \geq 64$, the losses are noticeable. The superior SE of M-MMSE for any value of M comes from the fact that it finds the optimal tradeoff between interference suppression and coherent combining of the desired signal. When compared with the UL sum SE 2.8 bit/s/Hz/cell achieved by basic LTE systems, the SE is increased by more than 10× with ZF when $M \geq 64$, which increases to 15× with M-MMSE. This provides evidence that the Massive MIMO technology is capable of improving the SE by an order of magnitude.

Table 2 reports the average UL sum SE for $M = 100$ and different pilot reuse factors. M-MMSE benefits particularly much from having $f > 1$. Thanks to the improved channel estimation quality, M-MMSE can better suppress the interference from UEs in the surrounding cells and the SE gain is increased. Since the other schemes do not suppress interference from other cells, their SE reduces when f is increased.

Table 2 Average UL sum SE for $M = 100$ and different pilot reuse factors f.

	$f = 1$	$f = 2$	$f = 4$
M-MMSE	48.71	54.91	56.09
S-MMSE	43.69	43.81	41.36
ZF	39.26	39.82	36.91
MR	18.94	18.42	16.55

4 Pilot contamination is not a fundamental asymptotic limitation

Marzetta showed in his seminal paper [7] that, under spatially uncorrelated channels, MR combining (and precoding) achieves a finite asymptotic SE as $M \to \infty$ whose value is only determined by pilot contamination. The same limit is achieved (but for a smaller number of antennas) by S-MMSE and ZF [20]. Since then, the following results about Massive MIMO had been taken for granted:

- Due to pilot contamination, the capacity saturates as $M \to \infty$;
- MR is asymptotically optimal;
- More sophisticated schemes than MR can only improve the SE for finite M.

Recently, [5, 6] showed that with M-MMSE and a tiny amount of spatial channel correlation, the capacity of Massive MIMO increases without bound in UL and DL as $M \to \infty$, even under pilot contamination. More precisely, [5] showed that, if the BS makes use of channel estimates from UEs in all cells, an unbounded capacity is achieved with Massive MIMO when the channel correlation matrices of the pilot-contaminating UEs are asymptotically linearly independent. This is generally the case in practice [16]. If also the diagonals of the correlation matrices are linearly independent, [5, 6] proved that it is sufficient to know these diagonals (not the full correlation matrices) to achieve an unbounded asymptotic SE. Similar results were shown in [21] for a generalized MR combining.

All this proves that the above results are not correct, and tells us how to design Massive MIMO in the years to come. Indeed, the purpose of analyzing the asymptotic SE when $M \to \infty$ is not to advocate BSs with a nearly infinite number of antennas—that is physically impossible in a finite-sized world and the conventional channel models will eventually break down since more power is received than was transmitted. The importance of asymptotics is instead what it tells us about practical networks as antennas become a commodity and are deployed everywhere. For example, consider a network with any finite number of UL pilot signals and of active UEs, each with a finite-valued SE requirement. The results in [5] imply that there exists a finite number of antennas, M, that allows to deliver any required SE even in the presence of pilot contamination. This is made possible by exploiting the spatial correlation that appears naturally in wireless channels. The lack of these insights has not prevented the first deployments of Massive MIMO, but will guide the evolution of the technology towards what we call Massive MIMO 2.0.

References

1. E. Björnson, J. Hoydis, and L. Sanguinetti (2017) "Massive MIMO networks: Spectral, energy, and hardware efficiency" *Foundations and Trends in Signal Processing* 3-4: 154 - 655.
2. J. G. Andrews and X. Zhang and G. D. Durgin and A. K. Gupta (2016) "Are we approaching the fundamental limits of wireless network densification?" *IEEE Commun. Mag.* 54:184 - 190.
3. T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, (2016) "Fundamentals of Massive MIMO" *Cambridge University Press*.
4. E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta (2014) "Massive MIMO for next generation wireless systems" *IEEE Commun. Mag.* 2: 186 - 195.
5. E. Björnson, J. Hoydis, and L. Sanguinetti (2018) "Massive MIMO has unlimited capacity" *IEEE Trans. Wireless Commun.* 1: 574 - 590.
6. E. Björnson, J. Hoydis, and L. Sanguinetti (2017) "Pilot contamination is not a fundamental asymptotic limitation in Massive MIMO" *IEEE International Conference on Communications (ICC)* 2017.
7. T. L. Marzetta (2010) "Noncooperative cellular wireless with unlimited numbers of base station antennas" *IEEE Trans. Wireless Commun.* 11: 3590 - 3600.
8. S. Parkvall and E. Dahlman and A. Furuskär and M. Frenne (2017) "NR: The New 5G Radio Access Technology" *IEEE Commun. Std. Mag.* 4: 24 - 30.
9. S. Malkowsky, J. Vieira, L. Liu, P. Harris, K. Nieman, N. Kundargi, I. C. Wong, F. Tufvesson, V. Owall, and O. Edfors (2017) "The worlds first real-time testbed for Massive MIMO: Design, implementation, and validation" *IEEE Access* 5: 9073 - 9088.
10. G. Liu, X. Hou, J. Jin, F. Wang, Q. Wang, Y. Hao, Y. Huang, X. Wang, X. Xiao, and A. Deng (2017) "3D-MIMO with massive antennas paves the way to 5G enhanced mobile broadband: From system design to field trials" *IEEE J. Sel. Areas Commun.* 6: 1222 - 1233.
11. A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire (2013) "Joint spatial division and multiplexing - the large-scale array regime" *IEEE Trans. Inf. Theory* 10: 6441 - 6463.
12. J. Choi, D. Love, and P. Bidigare (2014) "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory" *IEEE J. Sel. Topics Signal Process.* 5: 802 - 814.
13. X. Gao and O. Edfors and F. Rusek and F. Tufvesson (2015) "Massive MIMO Performance Evaluation Based on Measured Propagation Data" *IEEE Trans. Wireless Commun.* 7: 3899 - 3911.
14. K. Upadhya, S. A. Vorobyov, and M. Vehkapera (2017) "Superimposed pilots are superior for mitigating pilot contamination in Massive MIMO' *IEEE Trans. Signal Process.* 11: 2917 - 2932.
15. D. Verenzuela, E. Björnson,and L. Sanguinetti (2018) "Spectral and energy efficiency of superimposed pilots in uplink Massive MIMO" *IEEE Trans. Wireless Commun.* 11: 7099 - 7115.
16. X. Gao and O. Edfors and F. Tufvesson and E. G. Larsson (2015) "Massive MIMO in Real Propagation Environments: Do All Antennas Contribute Equally?" *IEEE Trans. Commun.* 11: 3917 - 3928.
17. J. Vieira and F. Rusek and O. Edfors and S. Malkowsky and L. Liu and F. Tufvesson (2017) "Reciprocity Calibration for Massive MIMO: Proposal, Modeling, and Validation" *IEEE Trans. Wireless Commun.* 5: 3042 - 3056.
18. H. Yin, D. Gesbert, M. Filippou, and Y. Liu (2013) "A coordinated approach to channel estimation in large-scale multiple-antenna systems" *IEEE J. Sel. Areas Commun.* 2: 264 - 273.
19. Xueru Li, Emil Björnson, Erik G. Larsson, Shidong Zhou, Jing Wang (2017) "Massive MIMO with Multi-cell MMSE Processing: Exploiting All Pilots for Interference Suppression" *EURASIP J. Wireless Commun. Networking* 117.
20. J. Hoydis, S. ten Brink, and M. Debbah (2013) "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.* 2: 160 - 171.
21. D. Neumann; T. Wiese, M. Joham, W. Utschick (2018) "A Bilinear Equalizer for Massive MIMO Systems" *IEEE Trans. Signal Process.* 14: 3740 - 3751.